

# Sparse Recovery of Positive Signals with Minimal Expansion

M.Amin Khajehnejad Alexandros G. Dimakis Weiyu Xu Babak Hassibi  
California Institute of Technology  
email: {amin,adim,weiyu,hassibi}@caltech.edu

February 24, 2009

## Abstract

We investigate the sparse recovery problem of reconstructing a high-dimensional non-negative sparse vector from lower dimensional linear measurements. While much work has focused on dense measurement matrices, sparse measurement schemes are crucial in applications, such as DNA microarrays and sensor networks, where dense measurements are not practically feasible. One possible construction uses the adjacency matrices of expander graphs, which often leads to recovery algorithms much more efficient than  $\ell_1$  minimization. However, to date, constructions based on expanders have required very high expansion coefficients which can potentially make the construction of such graphs difficult and the size of the recoverable sets small.

In this paper, we construct sparse measurement matrices for the recovery of non-negative vectors, using perturbations of the adjacency matrix of an expander graph with much smaller expansion coefficient. We present a necessary and sufficient condition for  $\ell_1$  optimization to successfully recover the unknown vector and obtain expressions for the recovery threshold. For certain classes of measurement matrices, this necessary and sufficient condition is further equivalent to the existence of a “unique” vector in the constraint set, which opens the door to alternative algorithms to  $\ell_1$  minimization. We further show that the minimal expansion we use is necessary for any graph for which sparse recovery is possible and that therefore our construction is tight. We finally present a novel recovery algorithm that exploits expansion and is much faster than  $\ell_1$  optimization. Finally, we demonstrate through theoretical bounds, as well as simulation, that our method is robust to noise and approximate sparsity.

# 1 Introduction

We investigate the problem of signal recovery in compressed sensing, i.e., the problem of reconstructing a signal  $\mathbf{x}$  that is assumed to be  $k$  sparse using  $m$  measurements,  $\mathbf{y} = \mathbf{Ax}$ , where  $m$  is smaller than the ambient dimension of the signal  $n$ , but larger than  $k$ .  $\mathbf{A}$  here is the  $m \times n$  so-called measurement matrix. In this paper, we focus on the case where the nonzero entries of  $\mathbf{x}$  are positive, a special case that is of great practical interest.

In compressed sensing,  $\mathbf{A}$  is often a dense matrix drawn from some ensemble of random matrices (see, e.g., [3]). In this paper, however, we will focus on sparse measurement matrices. This is important for numerous reasons. In several applications, like DNA micro arrays, the cost of each measurement increases with the number of coordinates of  $\mathbf{x}$  involved [16, 28]. Also, sparse measurement matrices often make possible the design of faster decoding algorithms (e.g., [11, 7, 8, 19]) apart from the general linear programming decoder [3]. In addition, unlike random measurement matrices (such as Gaussian or Bernoulli), which only guarantee the recovery of sparse vectors with high probability, expander graphs give deterministic guarantees (see, e.g., [11], which gives a deterministic guarantee for the fast algorithm proposed, and [6] for concentration lemmas on expander graphs).

Unlike Gaussian matrices, where reasonably sharp bounds on the thresholds which guarantee linear programming to recover sparse signals have been obtained [2], such sharp bounds do not exist for expander-graph-based measurements. This is the main focus of the current paper, for the special case where the  $k$ -sparse vector is non-negative.

It turns out that, due to the additional non-negativity constraint, one requires significantly fewer measurements to recover  $k$ -sparse non-negative signals. The non-negative case has also been studied in [5] for Gaussian matrices and also in the work of Bruckstein et al. [10], which further proposes a “matching pursuit” type of recovery algorithm. See also [29] for another example.

The success of a measurement matrix is often certified by a so-called Restricted Isometry Property (RIP) which guarantees the success of  $\ell_1$  minimization. Recently, Berinde et al. [6] showed that the adjacency matrices of suitable unbalanced expander graphs satisfy an RIP property for  $\ell_{p \sim 1}$  norm. However, it turns out that RIP conditions are only sufficient and often fail to characterize all good measurement matrices. A complete characterization of good measurement matrices was recently given in terms of their null space. More precisely, as stated in previous work (e.g. [17, 20, 22, 24]), if for any vector  $\mathbf{w}$  in the null space of  $\mathbf{A}$ , the sum of the absolute values of any  $k$  elements of  $\mathbf{w}$  is less than the sum of the absolute values of the rest of the elements, then the solution to  $\min \|\mathbf{x}\|_0$  subject to  $\mathbf{Ax} = \mathbf{y}$  can always be obtained by solving  $\min \|\mathbf{x}\|_1$  subject to  $\mathbf{Ax} = \mathbf{y}$ , provided  $\mathbf{x}$  is  $k$ -sparse.<sup>1</sup> This condition is stated in the work of Donoho [1] as the  $k$ -neighborly polytope property of  $\mathbf{A}$ , and in the work of Candes et al. as the uncertainty principle [3]. Donoho et al. also have been able to show the validity of this condition with high probability for random i.i.d Gaussian matrices and are therefore able to compute fairly tight thresholds on when linear-programming-based compressed sensing works [2]. The first analysis of the null space for expander graphs has been done by Indyk [9], where it was shown that every  $(2k, \epsilon)$  expander graph<sup>2</sup> with  $\epsilon \leq \frac{1}{6}$  will have a well supported null space. See also [18] for explicit constructions using expander graphs.

Furthermore, using Theorem 1 of [14], which is a generalization of the null space property

---

<sup>1</sup>Here  $\|\cdot\|_0$  represents the number of non-zero entries in its argument vector and  $\|\cdot\|_1$  is the standard  $\ell_1$ -norm.

<sup>2</sup>We shall formally define  $(k, \epsilon)$  expander graphs shortly.

theorem for the recovery of approximately sparse signals, Indyk’s result gives an upper bound on the error when linear programming is used to recover approximately  $k$ -sparse vectors using expander graph measurements.

**Contributions of the current paper.** We present a necessary and sufficient condition that completely characterizes the success of  $\ell_1$ -minimization *for non-negative signals*, similar to the null space condition for the general case. Our condition requires that all the vectors in the null space of  $\mathbf{A}$  have sufficiently large *negative support* (i.e. a large number of negative entries). It further turns out that, for a certain class of measurement matrices  $\mathbf{A}$ , this condition is nothing but the condition for the existence of a “unique” vector in the constraint set  $\{\mathbf{x}|\mathbf{Ax} = \mathbf{y}, \mathbf{x} \geq 0\}$ . This therefore suggests that any other convex optimization problem over this constraint set can find the solution. (We exploit this fact later to find faster alternatives to  $\ell_1$  minimization.)

We then use the necessary and sufficient characterization to construct sparse measurement matrices. Our construction relies on starting with the adjacency matrix of an unbalanced expander (with constant degree) and adding suitable small perturbations to the non-zero entries.

Several sparse matrix constructions rely on adjacency matrices of expander graphs [6, 9, 11, 13, 15]. In these works, the technical arguments require very large expansion coefficients, in particular,  $1 - \epsilon \geq 3/4$ , in order to guarantee a large number of unique neighbors [23] to the expanding sets. A critical innovation of our work is that we require much smaller expansion, namely,  $1 - \epsilon = 1/d$ , where  $d$  is the number of non-zero entries in every column of  $\mathbf{A}$ . In fact, we show that expansion of  $1 - \epsilon = \frac{1}{d}$  is *necessary* for any matrix that works for compressed sensing. These two results show that for nonnegative vectors, expansion of  $1 - \epsilon = \frac{1}{d}$  is necessary and sufficient, and the small expansion requirement allows a much larger set of recoverable signals.

The reason for this different requirement is that we use expansion in a different way than previous work; we do not require a unique neighbor property but rather rely on Hall’s theorem that guarantees that  $1 - \epsilon = 1/d$  expansion will guarantee perfect matchings for expanding sets. The matching combined with perturbations in the entries guarantees full rank sub-matrices which in turn translates to the null space characterization we need.

Finally, we propose a fast alternative to  $\ell_1$  optimization for recovering the unknown  $\mathbf{x}$ . The method first identifies a large portion of the unknown vector  $\mathbf{x}$  where the entries are zero, and then solves an “overdetermined” system of linear equations to determine the remaining unknown components of  $\mathbf{x}$ . Simulations are given to present the efficacy of the method and its robustness to noise.

## 2 Problem Formulation

The goal in compressed sensing is to recover a sparse vector from a set of under-determined linear equations. In many real world applications the original data vector is nonnegative, which is the case that we will focus on in this paper. The original problem of compressed sensing for the nonnegative input vectors is the following:

$$\min_{\mathbf{Ax}=\mathbf{y}, \mathbf{x} \geq 0} \|\mathbf{x}\|_0 \quad (1)$$

where  $\mathbf{A}^{m \times n}$  is the measurement matrix,  $\mathbf{y}^{m \times 1}$  is called the observation vector,  $\mathbf{x}^{n \times 1}$  is the unknown vector which is known to be  $k$ -sparse, i.e., to have only  $k$  nonzero entries, and where  $\|\cdot\|_0$  is  $l_0$  norm, i.e., the number of nonzero entries of a given vector. The typical situation is that  $n > m > k$ . Although (1) is an NP-hard problem, Donoho and Tanner have shown in [5] that, for a class of

matrices  $\mathbf{A}$  maintaining a so-called outwardly  $k$ -neighborly property and  $\mathbf{x}$  being at most  $k$ -sparse, the solution to (1) is unique and can be recovered via the following linear programming problem:

$$\min_{\mathbf{Ax}=\mathbf{y}, \mathbf{x} \geq 0} \|\mathbf{x}\|_1 \quad (2)$$

They also show that i.i.d Gaussian random  $m \times n$  matrices with  $m = n/2$  are outwardly  $m/8$ -neighborly with high probability, and thus allow the recovery of  $n/16$  sparse vectors  $\mathbf{x}$  via linear programming. They further define a *weak* neighborly notion, based upon which they show that the same Gaussian random matrices will allow the recovery of *almost* all  $0.279n$  sparse vectors  $\mathbf{x}$  via  $\ell_1$ -optimization for sufficiently large  $n$ .

In this paper, we primarily seek the answer to a similar question when the measurement matrix  $\mathbf{A}$  is sparse and, in particular when  $\mathbf{A}$  is the adjacency matrix of an unbalanced bipartite graph with constant left degree  $d$ . The aim is to analyze the outwardly neighborly conditions for this class of matrices and come up with sparse structures that allow the recovery of vectors with sparsity proportional to the number of equations.

### 3 Null Space Characterization and Complete Rank

We begin by stating an equivalent version of the outwardly neighborly condition which is in fact similar to the null space property that was mentioned in the introduction, but for the non-negative case. Later we show that this has a much more mundane interpretation for the special case of regular bipartite graphs. We present the first theorem in the same style as in [5].

**Theorem 3.1.** *Let  $\mathbf{A}$  be a nonnegative  $m \times n$  matrix and  $k < n/2$  be a positive integer. The following two statements are equivalent:*

1. *For every nonnegative vector  $\mathbf{x}_0$  with at most  $k$  nonzero entries,  $\mathbf{x}_0$  is the unique solution to (2) with  $\mathbf{y} = \mathbf{Ax}_0$ .*
2. *For every vector  $\mathbf{w} \neq 0$  in the null space of  $\mathbf{A}$ , and every index set  $S \subset \{1, 2, \dots, n\}$  with  $|S| = k$  such that  $\mathbf{w}_{S^c} \geq 0$ , it holds that*

$$\sum_{i=1}^n w_i > 0.$$

*Here  $S^c$  is the complement set of  $S$  in  $\{1, 2, \dots, n\}$  and  $\mathbf{w}_S$  denotes the sub-vector of  $\mathbf{w}$  constructed by those elements indexed in  $S$ .  $|S|$  means the cardinality of the set  $S$ .*

Theorem 3.1 is in fact the counterpart of Theorem 1 of [17] for nonnegative vectors. It gives a necessary and sufficient condition on the matrix  $\mathbf{A}$ , such that all  $k$ -sparse vectors  $\mathbf{x}_0$  can be recovered using (2). The condition is essentially that for every vector in the null space of  $\mathbf{A}$ , the sum of every  $n - k$  nonnegative elements should be greater than the absolute sum of the rest. (This is very similar, but not identical, to the null space property of [17].) Therefore we call it the non-negative null space property.

*Proof.* Suppose  $\mathbf{A}$  has the non-negative null space property. We assume  $\mathbf{x}_0$  is  $k$ -sparse and show that under the mentioned null space condition, the solution to (2) produces  $\mathbf{x}_0$ . We denote by  $\mathbf{x}_1$  the solution to (2). Let  $S$  be the support set of  $\mathbf{x}_0$ . We can write:

$$\begin{aligned}\|\mathbf{x}_1\|_1 &= \|\mathbf{x}_0 + (\mathbf{x}_1 - \mathbf{x}_0)\|_1 \\ &= \sum_{i=1}^n \mathbf{x}_0(i) + (\mathbf{x}_1 - \mathbf{x}_0)(i)\end{aligned}\tag{3}$$

$$= \|\mathbf{x}_0\|_1 + \sum_{i=1}^n (\mathbf{x}_1 - \mathbf{x}_0)(i)\tag{4}$$

Where  $\mathbf{x}_0(i)$  and  $(\mathbf{x}_1 - \mathbf{x}_0)(i)$  are the  $i$ th entry of  $\mathbf{x}_0$  and  $\mathbf{x}_1 - \mathbf{x}_0$  respectively. The reason (3) and (4) are true is that  $\mathbf{x}_1$  and  $\mathbf{x}_0$  are both nonnegative vectors and their  $\ell_1$ -norm is simply the sum of their entries. Now, if  $\mathbf{x}_1$  and  $\mathbf{x}_0$  are not equal, since  $\mathbf{x}_1 - \mathbf{x}_0$  is in the null space of  $\mathbf{A}$  and is non-negative on  $S_c$  (because  $S$  is the support set of  $\mathbf{x}_0$ ) we can write:

$$\sum_{i=1}^n (\mathbf{x}_1 - \mathbf{x}_0)(i) > 0\tag{5}$$

which implies

$$\|\mathbf{x}_1\|_1 > \|\mathbf{x}_0\|_1$$

But we know that  $\|\mathbf{x}_1\|_1 \leq \|\mathbf{x}_0\|_1$  from the construction. This means that we should have  $\mathbf{x}_1 = \mathbf{x}_0$ . Conversely, suppose there is a vector  $\mathbf{w}$  in the null space of  $\mathbf{A}$  and a subset  $S \subset \{1, 2, \dots, n\}$  of size  $k$  with  $\mathbf{w}_{S^c} \geq 0$  and  $\sum_{i=1}^n w_i \leq 0$ . We construct a non-negative vector  $\mathbf{x}_0$  supported on  $S$ , and show that there exist another nonnegative vector  $\mathbf{x}_1 \neq \mathbf{x}_0$  such that  $\mathbf{Ax}_0 = \mathbf{Ax}_1$  and  $\|\mathbf{x}_1\| \leq \|\mathbf{x}_0\|$ . This means that  $\mathbf{x}_0$  is not the unique solution to (2) with  $\mathbf{y} = \mathbf{Ax}_0$  and will complete the proof. For simplicity we may assume  $S = \{1, 2, \dots, k\}$ . We construct a nonnegative vector  $\mathbf{x}_0$  supported on  $S$  that cannot be recovered via  $\ell_1$ -minimization of (2). Wlog we write

$$\mathbf{w} = \begin{bmatrix} -\mathbf{w}_{S^-} \\ \mathbf{w}_{S^+} \\ \mathbf{w}_{S^c} \end{bmatrix},\tag{6}$$

where  $\mathbf{w}_{S^-}$  and  $\mathbf{w}_{S^+}$  are both nonnegative vectors. Now set

$$\mathbf{x}_0 = \begin{bmatrix} \mathbf{w}_{S^-} \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{x}_1 = \begin{bmatrix} 0 \\ \mathbf{w}_{S^+} \\ \mathbf{w}_{S^c} \end{bmatrix}.\tag{7}$$

■

In this paper we will be considering measurement matrices  $\mathbf{A}$  which possess the following two features: 1. the entries are non-negative and 2. the sum of the columns are constant. This class of matrices includes measurement matrices obtained from the adjacency matrix of regular left degree bipartite graphs (which have a constant number of ones in each column), as well as their perturbed versions introduced in section 3.3. For this class of matrices we actually show that the condition for the success of  $\ell_1$  recovery is simply the condition for there being a “unique” vector in the constraint set  $\{\mathbf{x} | \mathbf{Ax} = \mathbf{Ax}_0, \mathbf{x} \geq 0\}$ . To this end, we prove the following theorem

**Theorem 3.2.** Let  $A \in \mathcal{R}^{m \times n}$  be a matrix with non-negative entries and constant column sum. Then the following three statements are equivalent.

1. For all non-negative  $k$ -sparse  $\mathbf{x}_0$ , it holds that

$$\{\mathbf{x} | \mathbf{Ax} = \mathbf{Ax}_0, \mathbf{x} \geq 0\} = \{\mathbf{x}_0\}.$$

2. For every vector  $\mathbf{w} \neq 0$  in the null space of  $\mathbf{A}$ , and every index set  $S \subset \{1, 2, \dots, n\}$  with  $|S| = k$  such that  $\mathbf{w}_{S^c} \geq 0$ , it holds that

$$\sum_{i=1}^n w_i > 0.$$

3. For every subset  $S \subset \{1, 2, \dots, n\}$  with  $|S| = k$ , there exists no vector  $\mathbf{w} \neq 0$  in the null space of  $\mathbf{A}$  such that  $\mathbf{w}_{S^c} \geq 0$ .

Theorems 3.1 and 3.2 show that for the class of matrices with non-negative entries and constant column sum, the condition for the success of  $\ell_1$  recovery is simply the condition for there being a “unique” vector in the constraint set  $\{\mathbf{x} | \mathbf{Ax} = \mathbf{Ax}_0, \mathbf{x} \geq 0\}$ . In this case, *any* optimization problem, e.g.,  $\min_{\mathbf{x} \geq 0, \mathbf{Ax}=\mathbf{y}} \|\mathbf{x}\|_2$ , would also recover the desired  $\mathbf{x}_0$ .

In fact, rather than prove Theorem 3.2, we shall prove the following stronger result (from which Theorem 3.2 readily follows).

**Lemma 3.1.** Let  $A \in \mathcal{R}^{m \times n}$  be a matrix with non-negative entries and constant column sum. Then the following two statements are equivalent.

1. For all non-negative  $\mathbf{x}_0$  whose support is  $S$ , it holds that

$$\{\mathbf{x} | \mathbf{Ax} = \mathbf{Ax}_0, \mathbf{x} \geq 0\} = \{\mathbf{x}_0\}.$$

2. For every vector  $\mathbf{w} \neq 0$  in the null space of  $\mathbf{A}$  such that  $\mathbf{w}_{S^c} \geq 0$ , it holds that

$$\sum_{i=1}^n w_i > 0.$$

3. There exists no vector  $\mathbf{w} \neq 0$  in the null space of  $\mathbf{A}$  such that  $\mathbf{w}_{S^c} \geq 0$ .

*Proof.* First, we show that for any nonnegative matrix  $\mathbf{A}$ , statements 1 and 3 of Lemma 3.1 is equivalent. suppose that the condition 3 holds for a specific subset  $S \subset \{1, 2, \dots, n\}$ . Consider a nonnegative  $n \times 1$  vector  $\mathbf{x}_0$  supported on  $S$ . If there exists another nonnegative vector  $\mathbf{x}_1$  with the property that  $\mathbf{Ax}_1 = \mathbf{Ax}_0$ , then  $\mathbf{x}_1 - \mathbf{x}_0$  would be a vector in the null space of  $\mathbf{A}$  which is also nonnegative on  $S^c$ , due to the nonnegativity of  $\mathbf{x}_1$  and the fact that  $S$  is the support set of  $\mathbf{x}_0$ . This contradicts the earlier assumption of condition 2.

The proof of converse is also straight forward. Suppose the condition 1 holds for a specific subset  $S$  and all nonnegative vectors  $\mathbf{x}_0$  supported on  $S$ . Let's say one can find a nonzero vector  $\mathbf{w}$  in the null space of  $\mathbf{A}$  with  $\mathbf{w}_{S^c} \geq 0$ . As in the proof of Theorem 3.1, we may write  $\mathbf{w}$  as

$$\mathbf{w} = \begin{bmatrix} -\mathbf{w}_{S^-} \\ \mathbf{w}_{S^+} \\ \mathbf{w}_{S^c} \end{bmatrix} \quad (8)$$

where  $\mathbf{w}_{S^-}$  and  $\mathbf{w}_{S^+}$  are both nonnegative vectors. Now if

$$\mathbf{x}_0 = \begin{bmatrix} \mathbf{w}_{S^-} \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{x}_1 = \begin{bmatrix} 0 \\ \mathbf{w}_{S^+} \\ \mathbf{w}_{S^c} \end{bmatrix}, \quad (9)$$

then  $\mathbf{x}_0$  and  $\mathbf{x}_1$  are distinct nonzero vectors and belong to the set  $\{\mathbf{x} | \mathbf{Ax} = \mathbf{Ax}_0, \mathbf{x} \geq 0\}$ . This is a contradiction to the assumption we earlier made.

So far we have shown that for any nonnegative matrix  $\mathbf{A}$  the two statements 1 and 3 are equivalent. Now we show that for matrices with constant column sum the two statements 2 and 3 are equivalent. We make use of Lemma 3.2 in Section 3.1, that for this special class of matrices with constant column sum, every vector in the null space has a zero sum of entries. Therefore, statement 2 can be true only if there is no  $\mathbf{w}$  in the null space of  $\mathbf{A}$  with  $\mathbf{w}_{S^c} \geq 0$ . Conversely if the condition in statement 3 holds, then there is no  $\mathbf{w} \in \mathcal{N}(\mathbf{A}) \setminus \{0\}$  such that  $\mathbf{w}_{S^c}$  is nonnegative and therefore statement 2 is also true.

■

### 3.1 Null Space of Adjacency Matrices

As promised earlier, we will now assume that  $\mathbf{A}$  is the adjacency matrix of a bipartite graph with  $n$  nodes on the left and  $m$  nodes on the right. We also assume that the graph is left  $d$ -regular. In other words  $\mathbf{A}$  is a  $(m \times n)$  matrix with exactly  $d$  ones in each column. We will now give a series of results for such matrices. However, we should note that, unless stated otherwise, all these results continue to hold for the class of matrices with non-negative entries and constant column sum (the interested reader should be able to easily verify this).

**Lemma 3.2.** *Let  $\mathbf{A}^{m \times n}$  be a matrix with nonnegative entries and constant column sum. For any vector  $\mathbf{w}$  in the null space of  $\mathbf{A}$ , the following is true*

$$\sum_{i=1}^n w_i = 0 \quad (10)$$

*Proof.* Let  $\mathbf{1} = [1, 1, \dots, 1]$  be the  $m \times 1$  vector of all 1's. We have:

$$\mathbf{Aw} = 0 \Rightarrow \mathbf{1}\mathbf{Aw} = 0 \Rightarrow d \sum_{i=1}^n \mathbf{w}_i = 0 \quad (11)$$

■

**Theorem 3.3.** *For any matrix  $\mathbf{A}^{m \times n}$  with exactly  $d$  1's in each column and other entries zero, the following two statements are equivalent:*

- Every nonnegative vector  $\mathbf{x}_0$  with at most  $k$  nonzero entries is the unique solution to (2) with  $\mathbf{y} = \mathbf{Ax}_0$ .
- Every vector  $\mathbf{w}$  in the null space of  $\mathbf{A}$  has at least  $k + 1$  negative entries.

*Proof.* We only need to show that for any  $\mathbf{w} \in \mathcal{N}(\mathbf{A})/\{0\}$  the second statements of Theorem 3.1 and Theorem 3.3 are equivalent. Let's assume there exists a nonzero  $\mathbf{w} \in \mathcal{N}(\mathbf{A})$  with at most  $k$  negative entries. We use  $S_{\mathbf{w}}^+$ ,  $S_{\mathbf{w}}^-$  and  $S_{\mathbf{w}}^0$  to denote the support of positive, negative and zero

entries of  $\mathbf{w}$  respectively. By Lemma 3.2,  $\sum_{i=1}^n w_i = 0$ . Therefore if  $S = S_{\mathbf{w}}^+ \cup S_{\mathbf{w}}^0$ , then  $|S| \leq k$  and the nonnegative null space property is not satisfied (for the set  $S$ ).

The other direction is straightforward. If any  $\mathbf{w} \in \mathcal{N}(\mathbf{A})$  has  $k + 1$  negatives, there is no choice for  $S \subset \{1, 2, \dots, n\}$  of size  $k$ , with  $\mathbf{w}_{S^c} \geq 0$  and the nonnegative null space property is satisfied already. ■

These results show how the structure of the null space of the measurement matrix is related to the recoverability of sparse vectors. Thus to achieve our primary goal, which is constructing optimal sparse measurement matrices, we need to find bipartite graphs with non-negative null space properties up to a maximal sparsity (hopefully, proportional to the dimension  $n$ ). One promising choice would be the use of the adjacency matrix of expander graphs. However, rather than restrict ourselves to this choice, we present some theorems paraphrasing the null-space property and interpreting it in terms of other properties of matrices. This way we show that at some point expander graphs inevitably emerge as the best choice, and even further as a necessary condition for any measurement matrix.

### 3.2 Complete Rank and Natural Expansion

Before proceeding, let us consider the following two definitions, whose relation to the main topic will be shortly made apparent.

**Definition 1.** For a matrix  $\mathbf{A}^{m \times n}$  we define the Complete Rank of  $\mathbf{A}$  (denoted by  $\mathcal{Cr}(\mathbf{A})$ ) to be the maximum integer  $r_0$  with the property that every  $r_0$  columns of  $\mathbf{A}$  are linearly independent. In other words,  $\mathcal{Cr}(\mathbf{A}) = \min_{\mathbf{w} \in \mathcal{N}(\mathbf{A})} (|\text{Supp}(\mathbf{w})| - 1)$ , where by  $\text{Supp}(\mathbf{w})$  we mean the support set of  $\mathbf{w}$ .

This notion is also known in linear algebra as Kruskal rank (see [27]).

**Definition 2.** A left regular bipartite graph  $(X, Y, d)$  with  $X$  and  $Y$  the set of left and right vertices ( $|X| = n, |Y| = m$ ) and  $d$  the regular left degree is called a  $(k, \epsilon)$ -unbalanced expander if for every  $S \subset X$  with  $|S| \leq k$ , the following holds:  $|\Gamma(S)| \geq kd(1 - \epsilon)$ , where  $\Gamma(S)$  is the set of neighbors of  $S$  in  $Y$ .

The following lemma connects these two notions:

**Lemma 3.3.** Every bipartite graph with adjacency matrix  $\mathbf{A}$  and left degree  $d$  is a  $(\mathcal{Cr}(\mathbf{A}), \frac{d-1}{d})$  expander.

*Proof.* If  $S \subset X$  with  $|S| \leq \mathcal{Cr}(\mathbf{A})$  then the columns of  $\mathbf{A}$  corresponding to the elements of  $S$  are linearly independent. So the sub-matrix of  $\mathbf{A}$  produced by only those columns which correspond to  $S$  must be of full rank. Therefore, it must have at least  $|S|$  non-zeros rows, which is to say  $|\Gamma(S)| \geq |S| = |S|d(1 - \frac{d-1}{d})$ . ■

A direct corollary of this theorem is that:

$$\forall S \subseteq X, |\Gamma(S)| \geq \min(|S|, \mathcal{Cr}(\mathbf{A})) \quad (12)$$

The notion of complete rank is tightly related to the expansion property. It is also related to the null space characterization we are shooting for. The following theorem sheds some light on this fact.

**Theorem 3.4.** If  $\mathbf{A}^{m \times n}$  is the adjacency matrix of a left  $d$ -regular bipartite graph, then for every nonzero vector  $\mathbf{w}$  in the null space of  $\mathbf{A}$  the number of negative elements of  $\mathbf{w}$  is at least  $\frac{\mathcal{C}r(\mathbf{A})}{d}$ .

*Proof.* Let  $X$  and  $Y$  be the sets of left and right vertices of the bipartite graph corresponding to  $\mathbf{A}$  ( $X$  corresponds to columns of  $\mathbf{A}$ ). let  $S_{\mathbf{w}}^+$  be the set of vertices in  $X$  corresponding to the positive elements of  $\mathbf{w}$ , and likewise  $S_{\mathbf{w}}^-$  be set of vertices corresponding to the negative elements.<sup>3</sup> Let  $S_{\mathbf{w}} = S_{\mathbf{w}}^+ \cup S_{\mathbf{w}}^-$ . The equation  $\mathbf{Ax} = \mathbf{y}$  can be manifested on the graph representation of  $\mathbf{A}$  with each node of  $Y$  corresponding to an equation with zero R.H.S. This entails  $\Gamma(S_{\mathbf{w}}^+) = \Gamma(S_{\mathbf{w}}^-) = \Gamma(S_{\mathbf{w}})$ , since otherwise, there exists a vertex in  $Y$  connected to exactly one of the sets  $S_{\mathbf{w}}^+$  or  $S_{\mathbf{w}}^-$ , and its corresponding equation will not sum up to zero. On the other hand, from the definition of  $\mathcal{C}r(\mathbf{A})$ , we must have  $|S_{\mathbf{w}}| \geq \mathcal{C}r(\mathbf{A})$ . The number of edges emanating from  $S_{\mathbf{w}}^-$  is  $d|S_{\mathbf{w}}^-|$ , which is at least as large as the number of its neighbors  $|\Gamma(S_{\mathbf{w}}^-)|$ . Then:

$$d|S_{\mathbf{w}}^-| \geq |\Gamma(S_{\mathbf{w}}^-)| = |\Gamma(S_{\mathbf{w}})| \geq \mathcal{C}r(\mathbf{A})$$

Where the last inequality is a consequence of (12). ■

We now turn to the task of constructing adjacency matrices with complete rank proportional to dimension. Throughout this paper, all the thresholds that we achieve are asymptotic, i.e., they hold for the regime of very large  $n$  and  $m$ .

### 3.3 Perturbed Expanders

When  $n$  and  $m = \beta n$  are large, we are interested in constructing 0-1 matrices  $\mathbf{A}^{m \times n}$  with  $d$  (constant) 1's in each column such that  $\mathcal{C}r(\mathbf{A})$  is proportional to  $n$ . Furthermore, the maximum achievable value of  $\frac{\mathcal{C}r(\mathbf{A})}{nd}$  is critical. This is a very difficult question to address. However, it turns out to be much easier if we allow for a *small* perturbation of the nonzero entries of  $\mathbf{A}$ , as shown next.

**Lemma 3.4.** For a matrix  $\mathbf{A} \in \mathbb{R}^{m \times n}$  which is the adjacency matrix of a bipartite left  $d$ -regular graph, if the submatrix formed by any  $r_0$  columns of  $\mathbf{A}$  has at least  $r_0$  nonzero rows, then it is possible to perturb the nonzero entries of  $\mathbf{A}$  and obtain another nonnegative matrix  $\tilde{\mathbf{A}}$  through this procedure, with  $\mathcal{C}r(\tilde{\mathbf{A}}) \geq r_0$ . Furthermore, the perturbations can be done in a way that the sum of each column remains a constant  $d$ .

*Proof.* The proof is based on showing that the set of valid perturbations that do not guarantee  $\mathcal{C}r(\tilde{\mathbf{A}}) \geq r_0$  is measure zero. So, by choosing perturbations uniformly from the set of valid perturbations, with probability one we will have  $\mathcal{C}r(\tilde{\mathbf{A}}) \geq r_0$ . ■

**Remark 1.** It is worth mentioning that in a more practical scenario, what we really need is that every  $r_0$  columns of  $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{E}$  be not only nonsingular, but “sufficiently” nonsingular. In other words, if the minimum singular value of the submatrix formed by any  $r_0$  columns of  $\tilde{\mathbf{A}}$  is greater than a constant number  $c$  (which does not depend on  $n$ ), then that is what we recognize as the RIP – 2 condition for  $\tilde{\mathbf{A}}$ . This condition then guarantees that the solution to  $\ell_1$ -minimization is robust, in an  $\ell_2$ -norm sense, to the noise (please see [4] for more details on this issue). Besides, in order to

---

<sup>3</sup>We interchangeably use  $S$  and its variations to denote a set of vertices or a support set of a vector.

have a recovery algorithm which can be implemented with reasonable precision, it is important that we have such a well-conditioned statement on the measurement matrix. We will not delve into the details of this issue here. However, we conjecture that by leveraging ideas from perturbation theory, it is possible to show that  $\tilde{\mathbf{A}}$  maintains some sort RIP – 2 condition. Particularly, it has been shown in [26] that random dense perturbations  $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{E}$  will force the small singular values to increase, and the increment is proportional to  $\sqrt{m}$ . In our case where  $\mathbf{E}$  is a sparse matrix itself, we conjecture that the increment must be  $O(1)$ .

$\tilde{\mathbf{A}}$  corresponds to the same bipartite graph as  $\mathbf{A}$ . However, the edges are labeled with positive fractional weights between 0 and 1, rather than single 1 weights of  $\mathbf{A}$ . Besides, all the edges emanating from any node in  $X$  have a weight sum-up equal to  $d$ . It is worth noticing that, after modifying  $\mathbf{A}$  based on perturbations described above, Theorem 3.1, Lemma 3.2 and Theorems 3.3 and 3.4 all continue to hold for this class of matrices  $\tilde{\mathbf{A}}$ . Therefore  $\text{Cr}(\tilde{\mathbf{A}}) \geq r_0$  will guarantee perfect recovery of  $\frac{r_0}{d}$ -sparse vectors via  $\ell_1$ -minimization. Moreover, the fact that  $\text{Cr}(\tilde{\mathbf{A}}) \geq r_0$  can be translated back as if  $\mathbf{A}$  is a  $(r_0, \frac{d-1}{d})$  unbalanced expander graph. Therefore what we really care about is constructing  $(r_0, \frac{d-1}{d})$  expanders with  $\frac{r_0}{nd}$  as large as possible. In section 4, we use a probabilistic method to show that the desired  $(r_0 = \mu n, \frac{d-1}{d})$  expanders exist and give thresholds on  $\frac{\mu}{d}$ . Before continuing, note that we are using a  $1 - \epsilon = \frac{1}{d}$  expansion coefficient for perfect recovery, which is very small compared to other schemes that use expanders (see, e.g., [11, 6, 7, 9, 12, 13]) and require expansion coefficients at least larger than  $1 - \epsilon \geq \frac{3}{4}$ .  $1 - \epsilon = \frac{1}{d}$  is indeed the *critical* expansion coefficient. We shortly digress in a subsection to discuss this a little further.

### 3.4 Necessity of Expansion

Consider any  $m \times n$  measurement matrix  $A$  that allows the recovery of all  $r_0$ -sparse vectors and construct its corresponding bipartite graph  $B(n, m)$  by placing an edge between nodes  $i, j$  if the  $A_{ij}$  entry is nonzero. We show that any such matrix must correspond to an expander bipartite graph. The intuition is that a contracting set is certificate for a submatrix being rank-deficient, and hence reconstruction is impossible. In particular, for the case where each column of  $A$  has  $d$  non-zero entries we obtain the following statement:

**Lemma 3.5.** *Any  $m \times n$  measurement matrix  $\mathbf{A}$  with  $d$  non-zeros per column that allows the recovery of all  $r_0$ -sparse vectors must correspond to a  $(r_0, 1 - \frac{1}{d})$  bipartite expander graph.*

*Proof.* The statement holds for any recovery algorithm (even  $\ell_0$  minimization), because we show that even a stronger recovery algorithm that is given the support of the  $r_0$  sparse vector will fail to recover. Assume that the bipartite graph is not a  $(r_0, 1 - \frac{1}{d})$  expander, i.e. there exists a set of  $r \leq r_0$  columns that is adjacent to  $r - 1$  (or less) rows. Therefore the submatrix corresponding to these  $r$  columns must have rank strictly smaller than  $r$  regardless of what the non-zero entries are. By selecting a sparse signal that is supported exactly on these  $r$  columns we see that it is impossible for any algorithm to recover it, even if the support is known, since there is rank-loss in the corresponding measurement submatrix.

This argument is easily extended to the non-regular case where the number of non-zeros in every column is arbitrary. The key necessary condition is that every set of size  $r \leq r_0$  has a neighborhood of size  $r$  or greater. This is exactly the condition of Hall's marriage theorem that guarantees a perfect matching for every subset of size up to  $r_0$ . This perfect matching combined with perturbations will suffice to ensure that all the submatrices are full rank. ■

Gaussian and other dense measurement matrices correspond to completely connected bipartite graphs which obviously have the necessary expansion. Therefore, the previous lemma becomes interesting when one tries to construct sparse measurement matrices.

We can easily show that bigger expansion is also sufficient (but not necessary); given the adjacency matrix of a general  $(k, \epsilon)$ -expander, can we guarantee that with appropriate perturbations,  $\ell_1$ -optimization will allow the recovery of  $f(k, \epsilon)$ -sparse vectors for some positive function  $f(\cdot, \cdot)$ . The answer is yes, and the proof leverages on the fact that every  $(k, \epsilon)$  graph can be interpreted as a  $(k', 1 - \frac{1}{d})$ -expander for some  $k' \geq k$ .

**Lemma 3.6.** *If  $\epsilon \leq 1 - \frac{1}{d}$  and  $k > 0$ , then every  $d$ -left regular  $(k, \epsilon)$ -expander is also a  $(k(1 - \epsilon)d, 1 - \frac{1}{d})$ -expander*

*Proof.* Consider a subset  $S \subset X$  of the nodes with  $|S| \leq k(1 - \epsilon)d$ . If  $|S| \leq k$  then from the  $(k, \epsilon)$ -expansion  $|\Gamma(S)| \geq (1 - \epsilon)d|S| \geq |S|$ . If  $k < |S| \leq k(1 - \epsilon)d$ , then by choosing an arbitrary subset  $S' \subset S$  with  $|S'| = k$  and using the  $(k, \epsilon)$ -expansion on  $S'$  we have:

$$|\Gamma(S)| \geq |\Gamma(S')| \geq k(1 - \epsilon)d \geq |S|.$$

■

Recall that a  $(r_0, 1 - \frac{1}{d})$ -expander, when added with perturbations was capable of recovering  $r_0$ -sparse vectors. This and Lemma 3.6 immediately imply:

**Theorem 3.5.** *If  $\mathbf{A}$  is the adjacency matrix of an unbalanced  $d$ -left regular  $(k, \epsilon)$ -expander graph, then there exists a perturbation of  $\mathbf{A}$  in the nonzero entries resulting in a nonnegative matrix  $\tilde{\mathbf{A}}$ , such that every nonnegative  $k(1 - \epsilon)$ -sparse vector  $\mathbf{x}$  can be recovered from  $\mathbf{y} = \tilde{\mathbf{A}}\mathbf{x}$  without error using  $\ell_1$ -optimization.*

This is an improvement over the existing bounds. For example, [9], guarantees sparse  $\frac{k}{2}$  vectors can be reconstructed via  $\ell_1$ -optimization using  $(k, \epsilon)$  expanders with  $\epsilon \leq \frac{1}{6}$ . Using the above theorem with  $\epsilon = \frac{1}{6}$ ,  $\frac{5}{6}k$ -sparse non-negative vectors can be perfectly recovered by linear programming. Likewise [11] provides an algorithm that recovers vectors with sparsity  $\frac{k}{2}$  using  $(k, \epsilon)$  expander graphs with  $\epsilon \leq \frac{1}{4}$ . Our theorem for  $\epsilon = \frac{1}{4}$  allows for the recovery of  $\frac{3}{4}k$ -sparse positive signals. Note that these bounds are very conservative and in fact, the size of expanding sets when smaller expansion is required are much bigger, which yields much bigger provably recoverable sets.

## 4 Existence of Sparse Matrices with Linear Complete Rank

For fixed values of  $n > m > r_0$  and  $d$  we are interested in the question of the existence of  $(r_0, \epsilon = \frac{d-1}{d})$  expanders with constant left degree  $d$ . We use the standard first moment method argument to prove the existence of such an expander for appropriate ratios of  $n, m$  and  $r_0$  and  $d$ . The main result is given below, while the complete proof can be found in Appendix A.

**Theorem 4.1.** *For sufficiently large  $n$ , with  $m = \beta n$  and  $r_0 = \mu n$ , there exists a bipartite graph with left vertex size  $n$  and right size  $m$  which is a  $(r_0, \frac{d-1}{d})$  expander, if*

$$d > \frac{H(\mu) + \beta H(\frac{\mu}{\beta})}{\mu \log(\frac{\beta}{\mu})}. \quad (13)$$

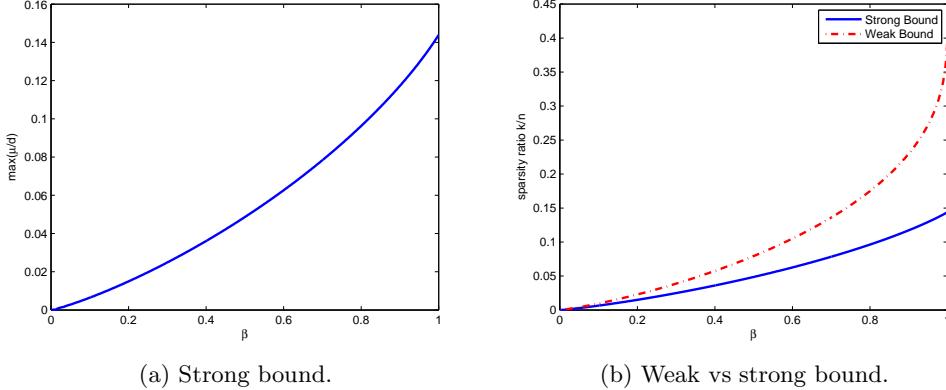


Figure 1: Comparison of weak achievable bound of Section 4.1 and with the strong achievable threshold of (13) for  $\frac{\mu}{d}$ .

More important is the question of how big the ratio  $\frac{\mu}{d}$  can be, since we earlier proved that we can recover up to  $\frac{r_0}{d} = \frac{\mu}{d}n$  sparse vectors using this method. Figure 1a illustrates the maximum achievable ratio for different values of  $\beta$  derived from (13).

#### 4.1 Weak bound

We are now interested in deriving conditions for recovering a specific support set  $S$  of size  $k = \alpha n$ , rather than obtain a worst case bound for matrices that work for all support sets. Recall that  $m = \beta n$ , left degree is  $d$ , and define  $\gamma_1 := (1 - e^{-d\frac{\alpha}{\beta}})\beta$ .

**Theorem 4.2.** *Define the function*

$$F(\rho_1, \rho_2) := \alpha H\left(\frac{\rho_1}{\alpha}\right) + (1 - \alpha)H\left(\frac{\rho_2}{1 - \alpha}\right) + \beta H\left(\frac{\rho_1 + \rho_2}{\beta}\right) + d(\rho_1 + \rho_2) \log\left(\frac{\rho_1 + \rho_2}{\beta}\right). \quad (14)$$

*For every  $\alpha$  such that  $F(\rho_1, \rho_2) < 0$  for every  $\rho_1, \rho_2$  that satisfies  $\rho_1 < \alpha, \rho_2 < 1 - \alpha, \rho_1 + \rho_2 < \gamma_1$ , a randomly selected subset of size  $k = \alpha n$  is recoverable from a random perturbed matrix  $\tilde{\mathbf{A}}$  with probability  $1 - o(1)$ .*

The bound that results from Theorem 4.2 is plotted in Figure 1b and has been compared to the strong threshold previously achieved. The full proof of this statement is given in Appendix B. The key argument is a matching condition for the recoverability of vectors supported on a specific subset  $S$ . The condition involves looking at the two-hop graph from  $S$  and checking if all sets of size up to  $|\Gamma(S)| + 1$  have a perfect matching:

**Lemma 4.1.** *Given a set  $S$  consider  $\Gamma(S)$  and denote  $S_2 = \Gamma(\Gamma(S)) \setminus S$ . Let the bipartite two-hop graph of  $S$  be denoted by  $B_S = (S \cup S_2, \Gamma(S \cup S_2))$ . Any non-negative vector  $\mathbf{x}_0$  supported on  $S$  can be recovered from  $\mathbf{y} = \tilde{\mathbf{A}}\mathbf{x}_0$ , if every subset  $S' \subset S \cup S_2$  of size  $|S'| \leq |\Gamma(S)| + 1$  has minimal expansion:  $|\Gamma(S')| \geq |S'|$ .*

*Proof.* Consider the two-hop bipartite graph of  $S$  and let  $C = (S \cup S_2)^c$  denote the remainder of the nodes in  $X$ . Further let  $\mathbf{A}_S$  denote the submatrix corresponding to  $B_S$ . By Hall's theorem

since every subset of  $B_S$  of size up to  $|\Gamma(S)| + 1$  has expansion equal to its size, it must also have a perfect matching and by the same perturbation argument  $\mathcal{Cr}(\tilde{A}_S) \geq |\Gamma(S)|$ .

By our null space characterization, to show that a set  $S$  can be recovered, it suffices to show that every nonzero vector  $\mathbf{w}$  in the nullspace of  $\tilde{A}$  cannot have all its negative components in  $S$ . Assume otherwise: that some  $\mathbf{w}$  has indeed all its negative support  $S_{\mathbf{w}}^- \subseteq S$ . Observe now that  $C$  cannot contain any of the positive support of  $\mathbf{w}$ , because every equation that is adjacent to a positive element must also be adjacent to a negative (since the matrix coefficients are positive) and  $\Gamma(S_{\mathbf{w}}^-)$  does not intersect  $\Gamma(C)$ . Therefore the whole support of  $\mathbf{w}$  must be contained in  $S \cup S_2$ .

Now we can show that  $|S_{\mathbf{w}}| \leq |\Gamma(S)|$ . Assume otherwise, that  $|S_{\mathbf{w}}| > |\Gamma(S)|$ . Then we could select a subset of  $K \subseteq S_{\mathbf{w}}$  such that  $|K| = |\Gamma(S)| + 1$ . This set  $K$  satisfies our assumption and is contained in  $B_S$  and therefore must have the minimal expansion  $|\Gamma(K)| \geq |K| = |\Gamma(S)| + 1$ . But since  $\Gamma(S_{\mathbf{w}}^-) = \Gamma(S_{\mathbf{w}}^+) = \Gamma(S_{\mathbf{w}})$  and  $K \subseteq S_{\mathbf{w}} \subseteq S$ , it must hold that  $|\Gamma(K)| \leq |\Gamma(S)|$ , which contradicts the minimal expansion inequality.

Therefore,  $|S_{\mathbf{w}}|$  must have a perfect matching which means that we can find a full rank submatrix  $\mathbf{A}_{\mathbf{w}}$  (corresponding to that matching) such that  $\mathbf{A}_{\mathbf{w}} \mathbf{w}_S = 0$  (where by  $\mathbf{w}_S$  we denote the vector  $\mathbf{w}$  restricted to its support). Since  $\mathbf{A}_{\mathbf{w}}$  is full rank,  $\mathbf{w}$  must be the all zeros vector which contradicts the assumption that  $S_{\mathbf{w}}^-$  can be contained in  $S$ . ■

## 5 Fast Algorithm

We now describe a fast algorithm for the recovery of sparse non-negative vectors from noiseless measurements. This algorithm relies on the minimal expansion we described in section 3.4. We employ a  $(kd + 1, 1 - \frac{1}{d})$  expander and perturb it as Lemma (3.4) to obtain a sparse nonnegative matrix  $\tilde{\mathbf{A}}$  with  $\mathcal{Cr}(\tilde{\mathbf{A}}) \geq kd + 1$ . Knowing that the target signal is at most  $k$ -sparse the algorithm works as follows

**Algorithm 1.** *Reverse Expansion Recovery*

1. Find zero entries of  $\mathbf{y}$  and denote them by  $\mathbf{y}_1$ . Also denote by  $T_1$  the index set of elements of  $\mathbf{y}_1$  in  $\mathbf{y}$ , and by  $T_2$  its complement. Wlog assume that  $\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}$ .
2. Locate in  $X$  the neighbors of the set of nodes in  $Y$  corresponding to  $T_1$ , name the set  $S_1$  and name the set of their complement nodes in  $X$  by  $S_2$ .
3. Identify the sub-matrix of  $\tilde{\mathbf{A}}$  that represents the nodes emanating from  $S_2$  to  $T_2$ . Call this sub-matrix  $\tilde{\mathbf{A}}_2$ . Columns of  $\tilde{\mathbf{A}}_2$  correspond to nodes in  $S_2$ , and its rows correspond to the nodes in  $T_2$ .
4. Set  $\hat{\mathbf{x}}_{S_1} = 0$  and compute  $\hat{\mathbf{x}}_{S_2} = \tilde{\mathbf{A}}_2^\dagger \mathbf{y}_2$ , where  $\mathbf{A}^\dagger$  is the pseudo-inverse of  $\mathbf{A}$  defined by  $\mathbf{A}^\dagger = (\mathbf{A}^t \mathbf{A})^{-1} \mathbf{A}^t$ . Declare  $\hat{\mathbf{x}}$  as the output.

The algorithm begins with identifying a big zero portion of the output and locating their corresponding nodes in  $Y$  (Refer to Figure 4 in Appendix C). In the next step, neighbors of these nodes are found in  $X$  and these two giant sets of nodes are eliminated from  $X$  and  $Y$ . Having done this, we are left with a much smaller system of linear equations, which turns out to be *over-determined*, and therefore our problem reduces to solving linear equations. The theoretical justification of the above statement, and why the algorithm works, is provided in Appendix C.

## 5.1 Noisy Case, Robust Algorithm

In general, the observations vector is contaminated with measurement noise, usually of significantly smaller power, leading to the following equation:

$$\mathbf{y} = \mathbf{Ax} + \mathbf{v}. \quad (15)$$

As before, it is assumed in (15) that  $\mathbf{x}$  is sparse.  $\mathbf{v}$  is the  $n \times 1$  observation noise vector with limited  $\ell_1$ -norm. In practice  $\mathbf{v}$  is often characterized by its  $\ell_2$  norm. However, in order to establish a recovery scheme that is robust to a limited power noise, we need to have a measurement matrix with a 2-RIP. This is not in general true for  $(0, 1)$  expander graphs, although it is realizable via a suitable perturbation for this class of matrices. However, for the scope of this paper, we assume that the limitation on the noise is given through its  $\ell_1$ -norm. This allows the use of the 1-RIP bounds of [6] to analyze the performance of our scheme in the presence of noise. Again we are assuming that  $\mathbf{x}$  is  $k$ -sparse

### Algorithm 2. Noisy Case

1. Sort elements of  $\mathbf{y}$  in terms of absolute value, pick the smallest  $m - kd$  ones and stack them in a vector denoted by  $\mathbf{y}_1$ . Also denote by  $T_1$  the index set of elements of  $\mathbf{y}_1$  in  $\mathbf{y}$ , and by  $T_2$  its complement. Wlog assume that  $\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}$ .
2. Locate in  $X$  the neighbors of the set of nodes in  $Y$  corresponding to  $T_1$ , name them by  $S_1$  and name the set of their complement nodes in  $X$  by  $S_2$ .
3. Identify the sub-matrix of  $\mathbf{A}$  that represents the nodes emanating from  $S_2$  to  $T_2$  referred to as  $\mathbf{A}_2$ . Columns of  $\mathbf{A}_2$  correspond to nodes in  $S_2$ , and its rows stand for the nodes in  $T_2$ .
4. Set  $\hat{\mathbf{x}}_{S_1} = 0$  and  $\hat{\mathbf{x}}_{S_2} = \operatorname{argmin}_{\mathbf{z} \in \mathbb{R}^{|S_2| \times 1}} \|\mathbf{A}_2 \mathbf{z} - \mathbf{y}_2\|_1$  and declare  $\hat{\mathbf{x}}$  as the output.

We show in the following theorem that our algorithm is robust (in a  $\ell_1$ -norm sense) to the observation noise.

**Theorem 5.1.** *If  $\mathbf{A}$  is the adjacency matrix of a  $(k, \epsilon)$ -expander with  $\epsilon < 0.5$ ,  $\mathbf{x}$  is a  $k$ -sparse nonnegative vector and  $\hat{\mathbf{x}}$  is the output of Algorithm 2, then  $\|\mathbf{x} - \hat{\mathbf{x}}\|_1 \leq \frac{6-4\epsilon}{1-2\epsilon} \|\mathbf{v}\|_1$*

*Proof.* Given in Appendix E. ■

## 6 Experimental Evaluation

We generated a random  $m \times n$  matrix  $\mathbf{A}$  with  $n = 2m = 500$ , and  $d = 3$  1's in each column. We then multiplied random sparse vectors with different sparsity levels by  $\mathbf{A}$ , and tried recovering them via the linear programming (2). Next, we added the perturbations described in section 3 to  $\mathbf{A}$  and applied the same sparse vectors to compare the recovery percentages in the two cases. This process was repeated for a few generations of  $\mathbf{A}$  and the best of the improvements we obtained is illustrated in Figure 2a.

In Figure 2b we plot the recovery percentage of Algorithm 1 for a random perturbed expander adjacency matrix  $\tilde{\mathbf{A}}$  of size  $250 \times 500$  and  $d = 6$ , and we have compared the performance with the

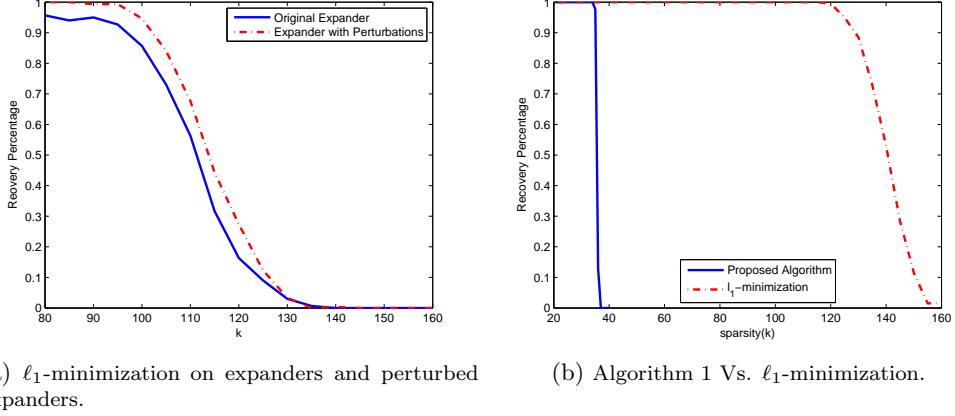


Figure 2: Percentage of successful recovery for  $\ell_1$ -minimization and Algorithm 1 with minimal expander measurements and reconstruction error of Algorithm 2.

$\ell_1$ -minimization method. Although the deterministic theoretical bounds of the two methods are the same, as observed in Figure 2b, in practice  $\ell_1$ -minimization is more effective for less sparse signals. However Algorithm 1 is considerably faster than linear programming and easier to implement.

In general, the complexity of Algorithm 1 is  $O(nk^2)$  which, when  $k$  is proportional to  $n$ , is similar to linear programming  $O(n^3)$ . However the constants are much smaller, which is of practical advantage. Furthermore, taking advantage of fast matrix inversion algorithms for very sparse matrices, Algorithm 1 can be performed in dramatically less operations. Figure 3 shows the Signal to Error Ratio as a function of Signal to Noise Ratio when Algorithm 2 (with  $\ell_2$ -norm used in step 4) has been used to recover noisy observations. Assuming that the output of the algorithm is  $\hat{\mathbf{x}}$ , Signal to Noise Ratio (SNR) and Signal to Error Ratio functions are defined as

$$SNR = 10 \log \frac{\|\mathbf{Ax}\|_2^2}{\|\mathbf{v}\|_2^2} \quad (16)$$

$$SER = 10 \log \frac{\|\mathbf{x}\|_2^2}{\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2} \quad (17)$$

Measurement matrices are the same as before.

## 7 Conclusion

In this paper we considered the recovery of a non-negative sparse vector using a sparse measurement matrix in the compressed sensing framework. We used the perturbed adjacency matrix of a bipartite expander graph to construct the sparse measurement matrix and proposed a novel fast algorithm. We computed recovery thresholds and showed that for measurement matrices with non-negative entries and constant column sum the constraint set  $\{\mathbf{x} | \mathbf{x} \geq 0, \mathbf{Ax} = \mathbf{y}\}$  is unique whenever  $\ell_1$  optimization is successful (which also means that any other optimization scheme would be successful). Finally, determining whether the matrices constructed satisfy an 2-RIP (we conjecture that they do), and constructing  $(0, 1)$  matrices that have complete rank proportional to  $n$  are open problems that may be worthy of further scrutiny.

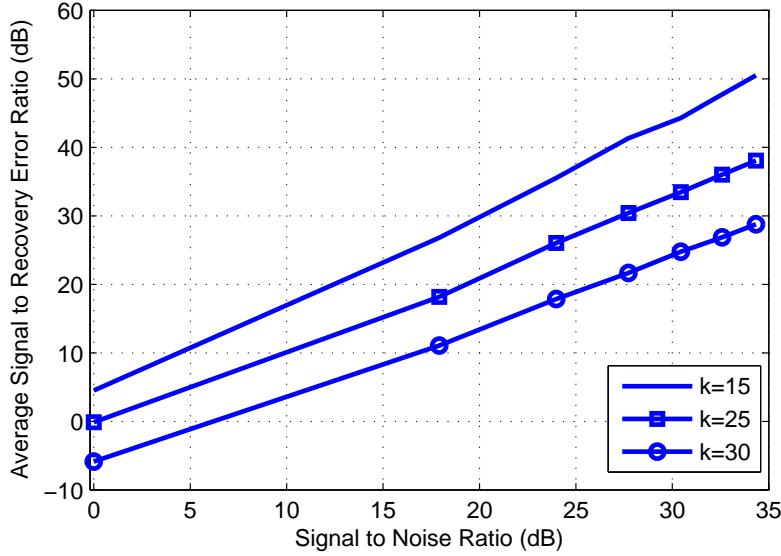


Figure 3: Simulation results for Algorithm 2; Signal to Error Ratio Vs. SNR.

## References

- [1] David Donoho, “High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension,” *Discrete and Computational Geometry*, 102(27), pp. 617-652 2006, Springer.
- [2] David Donoho and Jared Tanner “Neighborliness of randomly-projected simplices in high dimensions,” *Proc. National Academy of Sciences*, 102(27), pp. 9452-9457, 2005.
- [3] E.Candés, I.Romberg and T.Tao “Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information,” *IEEE Trans. on Info. Theory*, 52(2) pp. 489-509, Feb. 2006.
- [4] E. J. Candés. “The restricted isometry property and its implications for compressed sensing”, Compte Rendus de l’Academie des Sciences, Paris, Serie I, 346 589-592.
- [5] David L.Donoho and Jared Tanner “Sparse Nonnegative Solution of Underdetermined Linear Equations by Linear Programming”, *Proceedings of the National Academy of Sciences*, Vol. 102(27) (2005) 9446-9451.
- [6] R. Berinde, A. Gilbert, P. Indyk, H. Karloff, M. Strauss “Combining geometry and combinatorics: a unified approach to sparse signal recovery”, Forty-Sixth Annual Allerton Conference, September 23-26, 2008, University of Illinois at Urbana-Champaign, IL, USA.
- [7] P. Indyk and M. Ruzic “Near-Optimal Sparse Recovery in the  $L_1$  norm”, 49th Symposium on Foundations of Computer Science (FOCS), 2008.
- [8] J. Tropp and A. Gilbert, “Signal recovery from random measurements via orthogonal matching pursuit”. *IEEE Trans. on Information Theory*, 53(12) pp. 4655-4666, December 2007.
- [9] R. Berinde and P. Indyk “Sparse recovery using sparse matrices”, *Computer Science and Artificial Intelligence Laboratory Technical Report*, MIT-CSAIL-TR-2008-001,January 10, 2008.

- [10] Alfred M. Bruckstein, Michael Elad, and Michael Zibulevsky, “A non-negative and sparse enough solution of an underdetermined linear system of equations is unique”, Preprint, 2007.
- [11] Weiyu Xu and Babak Hassibi “Efficient compressive sensing with deterministic guarantees using expander graphs”, *IEEE Information Theory Workshop*, Lake Tahoe, September 2007.
- [12] Weiyu Xu and Babak Hassibi “Further Results on Performance Analysis for Compressive Sensing Using Expander Graphs”, Signals, Systems and Computers, 2007. ACSSC 2007. Conference Record of the Forty-First Asilomar Conference on 4-7 Nov. 2007 Page(s):621 - 625.
- [13] Sina Jafarpour, Weiyu Xu, Babak Hassibi, and Robert Calderbank “Efficient compressed sensing using high-quality expander graphs”. Preprint, 2008).
- [14] Weiyu Xu and Babak Hassibi “Compressed sensing over the Grassmann manifold: A unified analytical framework”. Preprint, 2008.
- [15] P. Indyk Explicit constructions for compressed sensing of sparse signals, SODA, 2008.
- [16] Farzad Parvaresh, Haris Vikalo, Sidhant Misra, and Babak Hassibi “Recovering sparse signals using sparse measurement matrices in compressed DNA microarrays” *IEEE Journal of Selected Topics In Signal Proc.*, Vol.2, Issue 3, pp. 275-285 06-2008.
- [17] Mihailo Stojnic, Weiyu Xu, and Babak Hassibi “Compressed sensing - probabilistic analysis of a null-space characterization” *IEEE International Conference on Acoustic, Speech and Signal Processing, ICASSP 2008*
- [18] V. Guruswami, J.R. Lee, A. Wigderson “Euclidean sections of  $L_1^n$  with sublinear randomness and error-correction over the reals”, RANDOM 2008.
- [19] V. Guruswami, J.R. Lee, A. Razborov “Almost Euclidean subspaces of  $L_1^n$  via expander codes”, SODA 2008.
- [20] A. Feuer and A. Nemirovski “On sparse representation in pairs of bases” *IEEE Transactions on Information Theory* 49(6): 1579-1581 (2003).
- [21] Michael Mitzenmacher, Eli Upfal “Probability and Computing: Randomized Algorithms and Probabilistic Analysis”. Cambridge University Press, 2005.
- [22] N. Linial and I. Novik “How neighborly can a centrally symmetric polytope be?” *Discrete and Computational Geometry*, 36(2006) 273-281.
- [23] M. Sipser and D.A. Spielman, “Expander Codes”, IEEE Transactions on Information Theory, 1996.
- [24] Y. Zhang “When is missing data recoverable”. Rice CAAM Department Technical Report TR05-15, 2005, available online at <http://www.dsp.ece.rice.edu/cs/>.
- [25] Michael R. Garey and David S. Johnson, “Computers and Intractability: A Guide to the Theory of NP-Completeness”, W.H. Freeman. ISBN 0-7167-1045-5.
- [26] G. W. Stewart “Perturbation Theory for the Singular Value Decomposition ”, Appeared in SVD and Signal Processing, II, R. J. Vacarro ed., Elsevier, Amsterdam, 1991.
- [27] J.B. Kruskal “Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics”, Linear Algebra Appl. 18 (1977), pp. 95138.

- [28] W. Dai, M. Sheikh, O. Milenkovic, and R. Baraniuk, “Compressive Sensing DNA Microarrays”, submitted, July 2008, avaialble online at <http://www-ece.rice.edu>.
- [29] W. Dai and O. Milenkovic, “Weighted Superimposed Codes and Integer Compressive Sensing”, submitted to IEEE Trans. on Inform. Theory, June 2008.

## A Proof of Theorem 4.1

Assuming that we generate a random matrix  $\mathbf{A}$  by randomly generating its columns, it suffices to show that the probability that  $\mathbf{A}$  has the desired expansion property is positive. For  $1 \leq i_1 < i_2 < \dots < i_r \leq n$  We denote by  $E_{i_1, i_2, \dots, i_r}$  the event that the columns of  $\mathbf{A}$  corresponding to the numbers  $i_1, i_2, \dots, i_r$  have at least  $n - r - 1$  entire 0 rows (rows that does not have a single non-zero elements in the columns  $\mathbf{A}_{i_1}, \mathbf{A}_{i_2}, \dots, \mathbf{A}_{i_r}$ ). In other words  $E_{i_1, i_2, \dots, i_r}$  is the event that the set of nodes  $\{i_1, i_2, \dots, i_r\}$  in  $X$  contracts in  $Y$ .

$$\begin{aligned} \mathbb{P}[\mathbf{A} \text{ is a } (r_0, \frac{d-1}{d})\text{-Exp.}] &= 1 - \mathbb{P}[\mathbf{A} \text{ not a } (r_0, \frac{d-1}{d})\text{-Exp.}] \\ &= 1 - \mathbb{P}\left[\bigcup_{d \leq r \leq r_0, 1 \leq i_1 < i_2 < \dots < i_r} E_{i_1, i_2, \dots, i_r}\right] \\ &= 1 - \sum_{r=d}^{r_0} \binom{n}{r} \mathbb{P}[E_{1, 2, \dots, r}] \end{aligned}$$

A combinatorial analysis yields the following:

$$\mathbb{P}[E_{1, 2, \dots, r}] \leq \frac{\binom{m}{r} \binom{r}{d}^r}{\binom{m}{d}^r}$$

Hence

$$\mathbb{P}[\mathbf{A} \text{ is a } (r_0, \frac{d-1}{d})\text{-Exp.}] \geq 1 - \sum_{r=d}^{r_0} \binom{n}{r} \frac{\binom{m}{r} \binom{r}{d}^r}{\binom{m}{d}^r} \quad (\text{A.18})$$

The objective is to show that this probability is positive. Equivalently, we show that for certain regimes of  $\beta, \mu$  and  $d$ , the summation on R.H.S of (A.18) vanishes. To this end, we split the sum into a sub-linear summation and a summation. We show that if  $d > 2$ , the sub-linear part will decays polynomially as  $n \rightarrow \infty$ , and the linear part decays exponentially in  $n$ , provided a certain relationship involving  $\beta, \mu$  and  $d$  holds. We state this in two different theorems.

**Theorem A.1.** *If  $0 < \alpha < e^{\frac{d}{2-d}} \beta^{\frac{1-d}{2-d}}$  and  $d \geq 3$  then  $\sum_{r=d}^{\alpha n} \binom{n}{r} \frac{\binom{m}{r} \binom{r}{d}^r}{\binom{m}{d}^r} = O(n^{1-d(d-2)})$ .*

*Proof.* We can write:

$$\sum_{r=d}^{\alpha n} \binom{n}{r} \frac{\binom{m}{r} \binom{r}{d}^r}{\binom{m}{d}^r} \leq \sum_{r=d}^{\alpha n} \binom{n}{r} \binom{m}{r} \left(\frac{r}{m}\right)^{rd} \quad (\text{A.19})$$

$$\leq \sum_{r=d}^{\alpha n} \left(\frac{ne}{r}\right)^r \left(\frac{me}{r}\right)^r \left(\frac{r}{m}\right)^{rd} \quad (\text{A.20})$$

$$= \sum_{r=d}^{\alpha n} \left(\frac{cr}{n}\right)^{r(d-2)} \quad (\text{A.21})$$

Where  $c = e^{\frac{2}{d-2}} \beta^{\frac{1-d}{d-2}}$ . (A.20) and (A.21) are deduced from the bounds  $\frac{\binom{r}{d}}{\binom{m}{d}^r} \leq \left(\frac{r}{m}\right)^d$  for  $r < m$ , and  $\binom{n}{k} \leq \left(\frac{ne}{k}\right)^k$  respectively. It is easy to show that when  $\alpha < \frac{1}{ec}$ ,  $\left(\frac{cr}{n}\right)^r$  is decreasing in  $r$ , and thus replacing all the terms in (A.21) by the first term will only increase the sum. The whole term is thus smaller than  $\alpha n \left(\frac{cd}{n}\right)^{d(d-2)} = \lambda n^{1-d(d-2)}$  for some positive constant  $\lambda$ . ■

**Theorem A.2.** For  $m = \beta n$  and  $r_0 = \mu n$ , if  $d > \frac{H(\mu) + \beta H(\frac{\mu}{\beta})}{\mu \log(\frac{\beta}{\mu})}$  then for any  $0 < \alpha < \mu$ , the sum  $\sum_{r=\alpha n+1}^{\mu n} \binom{n}{r} \frac{\binom{m}{r} \binom{r}{d}^r}{\binom{m}{d}^r}$  decays exponentially as  $n \rightarrow \infty$ .

*Proof.* Using the standard bounds of (F.49) on binomial coefficients we can write:

$$\sum_{r=\alpha n+1}^n \binom{n}{r} \frac{\binom{m}{r} \binom{r}{d}^r}{\binom{m}{d}^r} \leq n^2 \sum_{r=\alpha n+1}^{\mu n} 2^{nH(\frac{r}{n}) + mH(\frac{r}{m}) + r^2 H(\frac{d}{r}) - mrH(\frac{d}{m})} \quad (\text{A.22})$$

where  $H(x) = x \log(\frac{1}{x}) + (1-x) \log(\frac{1}{1-x})$  is the entropy function. Using  $H(\epsilon) = \epsilon \log(\frac{1}{\epsilon}) + O(\epsilon^2)$  for small  $\epsilon$  and the fact that as  $n \rightarrow \infty$   $\frac{d}{r} \rightarrow 0$  and  $\frac{d}{m} \rightarrow 0$  for  $\alpha n < r \leq \mu n$ , (F.49) can be written as:

$$\begin{aligned} \sum_{r=\alpha n+1}^n \binom{n}{r} \frac{\binom{m}{r} \binom{r}{d}^r}{\binom{m}{d}^r} &\leq n^2 \sum_{r=\alpha n+1}^{\mu n} 2^{nH(\frac{r}{n}) + mH(\frac{r}{m}) + rd \log_2 \frac{d}{m} + O(1)} \\ &= O(n^3 2^{n(H(\mu) + \beta H(\frac{\mu}{\beta}) + \mu d \log(\frac{\mu}{\beta}))}) \end{aligned} \quad (\text{A.23})$$

(A.23) vanishes if  $d > \frac{H(\mu) + \beta H(\frac{\mu}{\beta})}{\mu \log(\frac{\beta}{\mu})}$ . ■

The above argument leads to Theorem 4.1.

## B Derivation of the Weak Bound

We start with a straightforward modification of Theorem 3.1:

**Theorem B.1.** let  $\mathbf{x}_0 \in (\mathbb{R}^+)^n$  be fixed,  $\mathbf{y} = \mathbf{Ax}_0$  and denote by  $S$  the support of  $\mathbf{x}_0$ . Then the solution  $\mathbf{x}$  of (2) will be identical to  $\mathbf{x}_0$  if and only if there exists no  $\mathbf{w}$  in the null space of  $\mathbf{A}$  such that  $\mathbf{w}_{S^c}$  is nonnegative and  $\sum_{i=1}^n w_i > 0$ .

*Proof.* Similar to the proof of Theorem 3.1 with  $S$  as the support of  $\mathbf{x}_0$  ■

**Corollary 1.** *If  $\mathbf{A}$  is the adjacency matrix of a bipartite graph with left constant degree, and if  $\mathbf{x}_0$  is a fixed nonnegative vector and  $\mathbf{y} = \mathbf{Ax}_0$ , then the solution  $\mathbf{x}$  of (2) will be identical to  $\mathbf{x}_0$  if and only if there exists no  $\mathbf{w}$  in the null space of  $\mathbf{A}$  so that  $\mathbf{w}_S^c$  is a nonnegative vector, where  $S$  is the support set of  $\mathbf{x}_0$ . In other words  $\mathbf{x}_0$  will be recoverable via L.P from  $\mathbf{Ax}_0$  provided the support of  $\mathbf{x}_0$  does not include the index set of all negative elements of a vector in the null-space of  $\mathbf{A}$ .*

*Proof.* Directly from Theorem B.1 and Lemma 3.2. ■

This last statement allows us to derive a combinatorial matching condition for the recovery of a vector supported on a specific subset  $S$ . We repeat the statement of the lemma:

**Lemma B.1.** *Given a set  $S$  consider  $\Gamma(S)$  and denote  $S_2 = \Gamma(\Gamma(S)) \setminus S$ . Let the bipartite two-hop graph of  $S$  be denoted by  $B_S = (S \cup S_2, \Gamma(S \cup S_2))$ . Any non-negative vector  $\mathbf{x}_0$  supported on  $S$  can be recovered from  $\mathbf{y} = \tilde{\mathbf{A}}\mathbf{x}_0$ , if every subset  $S' \subset S \cup S_2$  of size  $|S'| \leq |\Gamma(S)| + 1$  has minimal expansion:  $|\Gamma(S')| \geq |S'|$ .*

Observe that the expectation is (asymptotically)  $\mathbb{E}\Gamma(S) = (1 - e^{-d\frac{|S|}{m}})\beta n =: \gamma_1 n$ . Using a standard Chernoff bound [21] it is easy to show that  $\Gamma(S)$  is concentrated around its expectation:

$$\mathbb{P}[\Gamma(S) \leq \mathbb{E}\Gamma(S) + \epsilon_1] > 1 - \frac{1}{n},$$

if  $|S| \geq c_1 n$ , for  $\epsilon_1 = c_2 \sqrt{n \log n}$ . Therefore we define the event  $E_1 = \{\Gamma(S) \leq \gamma_1 n + \epsilon_1\}$ .

Consider the random graph created from placing  $d$  non-zero entries (with repetition) in every column of  $\tilde{\mathbf{A}}$ . From the set  $S$ , form  $\Gamma(S)$ , the corresponding  $S_2$ , and finally the bipartite graph  $B_S = (S \cup S_2, \Gamma(S \cup S_2))$ . Using the given combinatorial condition, we can recover a signal supported on  $S$  if every subset  $S_i \subset S \cup S_2$  of size  $|r| \leq |\Gamma(S)| + 1$  has sufficient expansion:  $|\Gamma(S_i)| \geq r$  (note that subsequently we drop the  $+1$  term since it is negligible for large  $n$ ). First we condition on the concentration of  $\Gamma(S)$ :

$$\mathbb{P}[S \text{ not recoverable}] = \mathbb{P}[S \text{ not recoverable}|E_1]\mathbb{P}[E_1] + \mathbb{P}[S \text{ not recoverable}|E_1^c]\mathbb{P}[E_1^c] \quad (\text{B.24})$$

$$\leq \mathbb{P}[S \text{ not recoverable}|E_1](1 - \frac{1}{n}) + \frac{1}{n}, \quad (\text{B.25})$$

therefore it suffices to bound the probability conditioned on  $\Gamma(S)$  concentrated. We are going to do a union bound over all possible selections of  $r_1$  nodes in  $S$  and  $r_2$  nodes in  $S_2$  so that  $r_1 + r_2 \leq \Gamma(S) + \epsilon_1$ . Since we are conditioning on  $E_1$  it suffices to have  $r_1 + r_2 \leq \mathbb{E}\gamma_1 n$ . The second problem is that the set  $S_2$  is random and dependent on  $\Gamma(S)$ . We are going to avoid this conditioning by allowing the choice of  $r_2$  to range over all the nodes  $n - k$  nodes in  $S^c$ .

$$\mathbb{P}[S \text{ not recoverable}|E_1] \leq \sum_{r_1+r_2 \leq \gamma_1 n} \binom{k}{r_1} \binom{n-k}{r_2} \mathbb{P}(r_1, r_2 \text{ contracts}|E_1). \quad (\text{B.26})$$

Now the problem is that conditioning on  $E_1$  implies that the set  $r_1$  does not expand too much, so it is actually increasing the probability of the bad contraction event. We can however easily show that this increase is at most a factor of 2:

$$\mathbb{P}(r_1, r_2 \text{ contracts}|E_1) = \frac{\mathbb{P}(r_1, r_2 \text{ contracts} \cap E_1)}{\mathbb{P}(E_1)} \leq \frac{\mathbb{P}(r_1, r_2 \text{ contracts})}{\mathbb{P}(E_1)}. \quad (\text{B.27})$$

Now since  $\mathbb{P}(E_1) \geq 1 - 1/n$ , for sufficiently large  $n$ ,  $1/\mathbb{P}(E_1) \leq 2$ , so

$$\mathbb{P}(r1, r2 \text{ contracts}|E_1) \leq 2\mathbb{P}(r1, r2 \text{ contracts}). \quad (\text{B.28})$$

The probability that the set  $r1, r2$  contracts can be further bounded by assuming  $\Gamma(r1, r2) = r1 + r2$  (any smaller neighborhood will have smaller probability) so

$$\mathbb{P}(r1, r2 \text{ contracts}) \leq \binom{m}{r1 + r2} \left(\frac{r1 + r2}{m}\right)^{d(r1 + r2)}.$$

Putting everything together we obtain the bound

$$\mathbb{P}[S \text{ not recoverable}|E_1] \leq 2 \sum_{r1 + r2 \leq \gamma_1 n} \binom{k}{r1} \binom{n - k}{r2} \binom{m}{r1 + r2} \left(\frac{r1 + r2}{m}\right)^{d(r1 + r2)}. \quad (\text{B.29})$$

We move everything to the exponent and use standard binomial approximations to obtain

$$\mathbb{P}[S \text{ not recoverable}|E_1] \leq 2 \sum_{r1 + r2 \leq \gamma_1 n} 2^{kH(\frac{r1}{k}) + (n-k)H(\frac{r2}{n-k}) + mH(\frac{r1+r2}{m}) + d(r1+r2)\log(\frac{r1+r2}{m})}. \quad (\text{B.30})$$

Recall that the recoverable fraction is  $k = \alpha n$ ,  $m = \beta n$ , and denote  $\rho_1 = r_1/n$ ,  $\rho_2 = r_2/n$ . Define the function

$$F(\rho_1, \rho_2) := \alpha H\left(\frac{\rho_1}{\alpha}\right) + (1 - \alpha)H\left(\frac{\rho_2}{1 - \alpha}\right) + \beta H\left(\frac{\rho_1 + \rho_2}{\beta}\right) + d(\rho_1 + \rho_2)\log\left(\frac{\rho_1 + \rho_2}{\beta}\right), \quad (\text{B.31})$$

and observe that the bound on the probability of failure (B.30) becomes

$$\mathbb{P}[S \text{ not recoverable}|E_1] \leq 2 \sum_{r1 + r2 \leq \gamma_1 n} 2^{nF(\rho_1, \rho_2)}.$$

Therefore for a fixed  $\beta, \gamma_1$  we are trying to find the largest  $\alpha^*$  that makes  $F(\rho_1, \rho_2)$  negative for every  $\rho_1, \rho_2$  for which  $\rho_1 + \rho_2 \leq \gamma_1$ . For this  $\alpha^*$ , we can recover with exponentially high probability conditioned on the sublinear sets do not contract (which has been already established).

## C Proof of the Validity of Algorithm 1

As mentioned earlier and illustrated in Figure 4, the algorithm begins by identifying a big zero portion of the output and eliminating two large sets of nodes from  $X$  and  $Y$ . Having done this, we are left with a much smaller system of linear equation, which turns out to be an over-determined system and can be uniquely solved using matrix inversions. The theoretical justification of the above statement, and why the algorithm works, is provided in Appendix C. The fact that it is over-determined is secured by the expansion property of the measurement matrix and its proportional complete rank. Figure 4 is a graphical illustration of how Algorithm1 decomposes  $X$  and  $Y$  into subsets and makes the search and observation spaces shrink from  $X$  and  $Y$  into  $S_2$  and  $T_2$  respectively. Intuitively speaking, when we isolate a big proportion of the nodes on the right, their neighbors in  $X$  are big enough to leave us with a set of nodes on the left that are less than the

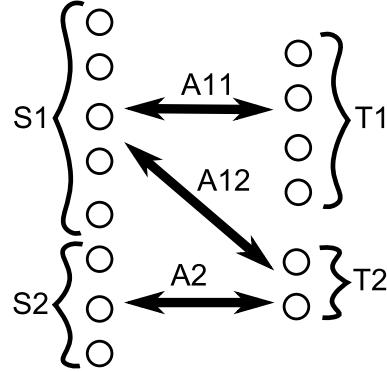


Figure 4: Decomposition of nodes and edges by Algorithm 1.

remaining nodes on the right. This procedure is therefore nothing but a block diagonalization (by rearranging rows and columns) of  $\tilde{\mathbf{A}}$  into a lower triangular matrix:

$$\begin{bmatrix} \tilde{\mathbf{A}}_{11} & 0 \\ \tilde{\mathbf{A}}_{12} & \tilde{\mathbf{A}}_2 \end{bmatrix} \quad (\text{C.32})$$

where  $\tilde{\mathbf{A}}_2$  is a square or tall full rank matrix. The following theorem certifies that Algorithm 1 is indeed valid and it recovers any  $k$ -sparse vector without error.

**Theorem C.1. Validity of Algorithm**

If  $\mathbf{x}$  is a  $k$ -sparse non-negative vector and  $\tilde{\mathbf{A}}$  is a perturbed  $(kd+1, 1 - \frac{1}{d})$  expander with  $\mathcal{Cr}(\tilde{\mathbf{A}}) \geq kd+1$  then:

1.  $\mathbf{y}$  is  $kd$ -sparse and therefore has at least  $m - kd$  zeros.
2.  $|S_2| \leq |T_2|$  and  $\tilde{\mathbf{A}}_2$  is full rank.
3.  $\hat{\mathbf{x}} = \mathbf{x}$

*Proof.*

1. Trivial by noting the fact that the graph representation of  $\tilde{\mathbf{A}}$  is left  $d$ -regular.
2. Suppose  $|S_2| > |T_2| = kd$ . select an arbitrary subset of  $S_2' \subseteq S_2$  of size  $kd+1$ . Because of the expansion property:  $|\Gamma(S_2')| \geq kd+1 > |T_2|$ . But  $\Gamma(S_2')$  is in  $T_2$  and this is a contradiction. Diagonalization of (C.32) and the fact that  $\tilde{\mathbf{A}}_2$  is a tall matrix and  $\mathcal{Cr}(\tilde{\mathbf{A}}) \geq kd+1$  together imply that  $\tilde{\mathbf{A}}_2$  has full column rank.
3. Every node in  $Y$  is an equation with nonnegative variables from  $\mathbf{x}$ , and positive weights from edges of the graph. If any entry in  $\mathbf{x}_{S_1}$  is greater than zero, then the equations corresponding to its neighbors in  $T_1$  are not zero (and there is at least one such equation since  $S_1 = \Gamma(T_1)$ ). This is in contradiction with the choice of  $T_1$ . Therefore  $\mathbf{x}_{S_1} = 0 = \hat{\mathbf{x}}_{S_1}$ . Also since  $\tilde{\mathbf{A}}_2 \mathbf{x}_{S_2} = \mathbf{y}_2$  and  $\tilde{\mathbf{A}}_2 \hat{\mathbf{x}}_{S_2} = \mathbf{y}_2$  and  $\tilde{\mathbf{A}}_2$  is full rank we conclude that  $\mathbf{x}_{S_2} = \hat{\mathbf{x}}_{S_2}$ .

■

**Remark 2.** Note that in proving the last part of Theorem C.1, we used the fact that  $\mathbf{x}$  is a non-negative vector. This proof does not hold for general sparse vectors.

**Remark 3.** By proving the part 2 of C.1 we implicitly proved that every expander from  $X$  to  $Y$  is a contractor from  $Y$  to  $X$ . This can be generalized as following :

$$\forall T \subset Y, |T| \leq rd + 1 \Rightarrow |\Gamma(T)| \leq |T|. \quad (\text{C.33})$$

## D 1-RIP for Expanders

We present a simple argument that the perturbed matrix  $\tilde{\mathbf{A}}$  has the 1-RIP property if the underlying graph is an  $(k, \epsilon)$ , for  $\epsilon < 1/2$ . The 1-RIP property states that for every  $k$ -sparse vector  $\mathbf{x}$  and suitable constants  $c_1, c_2$ , the  $\ell_1$  norm of  $\|\mathbf{Ax}\|_1$  is close to the norm of  $\mathbf{x}$ :

$$(1 - c_1)\|\mathbf{x}\|_1 \leq \|\mathbf{Ax}\|_1 \leq (1 + c_2)\|\mathbf{x}\|_1. \quad (\text{D.34})$$

Berinde et al. [6] already show that adjacency matrices of expander graphs will have this property, also for  $p$  norms where  $p \leq 1 - 1/\log n$ . The argument we present here also requires  $\epsilon < 1/2$ , but is arguably simpler and easily extends to the perturbed case.

Consider  $\tilde{\mathbf{A}}$  to be the perturbed adjacency matrix of a  $(k, \epsilon)$  unbalanced expander for  $\epsilon < 1/2$  and each nonzero entry is in  $[1 - \epsilon_1, 1 + \epsilon_1]$ . Consider  $S$ , the support set of  $\mathbf{x}$ . By Hall's theorem since every set  $S$  of size up to  $k$  has  $d(1 - \epsilon)|S|$  neighbors, there must exist a  $d(1 - \epsilon)$ -matching, i.e. every node in  $S$  can be matched to  $d(1 - \epsilon)$  left nodes. Decompose the measurement matrix

$$\tilde{\mathbf{A}} = \mathbf{A}_M + \mathbf{A}_C. \quad (\text{D.35})$$

Where  $\mathbf{A}_M$  is supported on the  $d(1 - \epsilon)$ -matching (i.e every row has one non-zero entry and every column has  $d(1 - \epsilon)$  non-zero entries). The remainder matrix  $\mathbf{A}_C$  has  $\epsilon d$  non-zero entries in each column, and notice that the decomposition is adapted to the support of the vector  $\mathbf{x}$ . By the triangle inequality:

$$\|\mathbf{Ax}\|_1 \geq \|\mathbf{A}_M\mathbf{x}\|_1 - \|\mathbf{A}_C\mathbf{x}\|_1. \quad (\text{D.36})$$

It is easy to see that

$$\|\mathbf{A}_M\mathbf{x}\|_1 \geq (1 - \epsilon_1)d(1 - \epsilon)\|\mathbf{x}\|_1, \quad (\text{D.37})$$

since  $\mathbf{A}_M\mathbf{x}$  is a vector that contains  $d(1 - \epsilon)$  copies of each entry of  $\mathbf{x}$  multiplied by coefficients in  $[1 - \epsilon_1, 1 + \epsilon_1]$ . Also since each column of  $\|\mathbf{A}_C\|_1$  contains  $\epsilon d$  non-zero entries,

$$\|\mathbf{A}_C\mathbf{x}\|_1 \leq (1 + \epsilon_1)\epsilon d\|\mathbf{x}\|_1, \quad (\text{D.38})$$

since each entry of  $\mathbf{A}_M\mathbf{x}$  is a summation of terms of  $\mathbf{x}$  and  $\|\mathbf{A}_M\mathbf{x}\|$  is also a summation in which each entry of  $\mathbf{x}$  appears  $d\epsilon$  times, multiplied by coefficients in  $[1 - \epsilon_1, 1 + \epsilon_1]$ . The same argument yields the upper bound:

$$\|\mathbf{Ax}\|_1 \leq d\|\mathbf{x}\|_1 \quad (\text{D.39})$$

Therefore, putting these together we obtain:

$$(1 - \epsilon_1 - 2\epsilon)\|\mathbf{x}\|_1 \leq \|\mathbf{Ax}\|_1 \leq d\|\mathbf{x}\|_1. \quad (\text{D.40})$$

Therefore we need an expander with  $\epsilon \geq (1 - \epsilon_1)/2$  which for arbitrarily small perturbations goes to the  $\epsilon < 1/2$  limit.

## E Proof of Robustness of Algorithm 2

We first state the following lemma from [6]:

**Lemma E.1.** *Consequence of Lemma 9 of [6]: If  $A$  is the adjacency matrix of a  $(k,\epsilon)$ -expander with  $\epsilon < 0.5$  and  $\mathbf{u}$  is a  $k$ -sparse vector, then  $d(1 - 2\epsilon)\|\mathbf{u}\|_1 \leq \|\mathbf{Au}\| \leq d\|\mathbf{u}\|_1$ .*

We have presented a new proof of this Lemma in the Appendix D, which is based on Generalized Hall's Theorem. That proof allows us to state with certainty that with suitable perturbations,  $\tilde{\mathbf{A}}$  will also have a 1-RIP.

By rearranging the rows and columns of  $\mathbf{A}$ , we may assume  $\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix}$ ,  $\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}$ ,  $\mathbf{v} = \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \end{bmatrix}$  and  $\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & 0 \\ \mathbf{A}_{12} & \mathbf{A}_2 \end{bmatrix}$ , where  $\mathbf{y}_1$  and  $\mathbf{y}_2$  are those obtained by the algorithm,  $\mathbf{x}_1 = \mathbf{x}_{S_1}$  and  $\mathbf{x}_2 = \mathbf{x}_{S_2}$ . Also let  $\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}}$  be the reconstruction error vector. By (15) we then have

$$\mathbf{y}_1 = \mathbf{A}_{11}\mathbf{x}_1 + \mathbf{v}_1 \quad (\text{E.41})$$

$$\mathbf{y}_2 = \mathbf{A}_{12}\mathbf{x}_1 + \mathbf{A}_2\mathbf{x}_2 + \mathbf{v}_2 \quad (\text{E.42})$$

$$\mathbf{e} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 - \hat{\mathbf{x}}_2 \end{bmatrix} \quad (\text{E.43})$$

Hence we have:

$$\|\mathbf{x}_1\|_1 \leq \|\mathbf{A}_{11}\mathbf{x}_1\|_1 = \|\mathbf{y}_1 - \mathbf{v}_1\|_1 \leq \|\mathbf{y}_1\|_1 + \|\mathbf{v}_1\|_1 \leq 2\|\mathbf{v}_1\|_1 \quad (\text{E.44})$$

The first inequality holds as a result of nonnegativity of  $\mathbf{x}_1$  and  $\mathbf{A}_{11}$ , and the fact that every column of  $\mathbf{A}_{11}$  has at least 1 nonzero entry. The last inequality is a consequence of the choice of  $\mathbf{y}_1$  in step 1 of the algorithm and the fact that  $\mathbf{Ax}$  is  $m - rd$  sparse. Let's assume  $\mathbf{y}_2 = \mathbf{A}_2\mathbf{x}_2 + \delta_2$ . From the way  $\hat{\mathbf{x}}_2$  is driven in step 4 of the algorithm it follows that:

$$\|\delta_2\|_1 \leq \|\mathbf{A}_{12}\mathbf{x}_1 + \mathbf{v}_2\|_1 \quad (\text{E.45})$$

And thus

$$\|\mathbf{A}_2(\mathbf{x}_2 - \hat{\mathbf{x}}_2)\|_1 = \|\delta_2 - \mathbf{A}_{12}\mathbf{x}_1 - \mathbf{v}_2\|_1 \leq 2\|\mathbf{A}_{12}\mathbf{x}_1 + \mathbf{v}_2\|_1 \leq 2d\|\mathbf{x}_1\|_1 + 2\|\mathbf{v}_2\|_1 \quad (\text{E.46})$$

Using this along with the 1-RIP condition of Lemma E.1 for the sparse vector  $\mathbf{u} = \begin{bmatrix} 0 \\ \mathbf{x}_2 - \hat{\mathbf{x}}_2 \end{bmatrix}$  we get:

$$c_1\|\mathbf{x}_2 - \hat{\mathbf{x}}_2\|_1 \leq 2d\|\mathbf{x}_1\|_1 + 2\|\mathbf{v}_2\|_1 \quad (\text{E.47})$$

Where  $c_1 = (1 - 2\epsilon)d$ . Equations (E.44) and (E.47) result in:

$$\|\mathbf{e}\|_1 \leq (2 + \frac{4d}{c_1})\|\mathbf{v}_1\|_1 + \frac{2}{c_1}\|\mathbf{v}_2\|_1 \leq \frac{6 - 4\epsilon}{1 - 2\epsilon}\|\mathbf{v}\|_1 \quad (\text{E.48})$$

Therefore we have been able to bound the  $\ell_1$  norm of the error with a constant factor of  $\ell_1$  norm of noise as desired.

**Remark 4.** As soon as a  $p$ -RIP can be proven to hold for  $\mathbf{A}$ , a similar statement bounding  $\ell_p$  norm of error with that of noise can be authentically established. However step 4 of the algorithm needs to be revised and  $\ell_1$  optimization should be replaced with  $\ell_p$ . In particular  $p = 2$  is of practical interest and the  $\ell_2$  optimization of step 4 is equivalent to the pseudo-inverse multiplication of the noise-less algorithm. However, as we mentioned earlier 2-RIP does not hold for 0-1 matrices. We speculate that a suitable perturbation can force singular values of  $\mathbf{A}$  to jump above a constant, and thus gift  $\mathbf{A}$  a 2-RIP condition.

## F Elementary bounds on binomial coefficients

For each  $\beta \in (0, 1)$ , define the binomial entropy  $H(\beta) = -\beta \log_2 \beta - (1 - \beta) \log_2(1 - \beta)$  (and  $H(0) = H(1) = 0$  by continuity). We make use of the following standard bounds on the binomial coefficients

$$n \left[ H\left(\frac{k}{n}\right) - \frac{\log_2(n+1)}{n} \right] \leq \log_2 \binom{n}{k} \leq n \left[ H\left(\frac{k}{n}\right) + \frac{\log_2(n+1)}{n} \right]. \quad (\text{F.49})$$